
Relevance of Motion-Related Assessment Metrics in Laparoscopic Surgery

Ignacio Oropesa, Magdalena K. Chmarra, Patricia Sánchez-González, Pablo Lamata, Sharon P. Rodrigues, Silvia Enciso, Francisco M. Sánchez-Margallo, Frank-Willem Jansen, Jenny Dankelman and Enrique J. Gómez

Relevance of Motion-Related Assessment Metrics in Laparoscopic Surgery

Ignacio Oropesa, MSc , Magdalena K. Chmarra, PhD ,
Patricia Sánchez-González, PhD , Pablo Lamata, PhD ,
Sharon P. Rodrigues, MD , Silvia Enciso, DVM ,
Francisco M. Sánchez-Margallo, PhD , Frank-Willem Jansen, MD, PhD ,
Jenny Dankelman, PhD , and Enrique J. Gómez, PhD

Abstract

Introduction. Motion metrics have become an important source of information when addressing the assessment of surgical expertise. However, their direct relationship with the different surgical skills has not been fully explored. The purpose of this study is to investigate the relevance of motion-related metrics in the evaluation processes of basic psychomotor laparoscopic skills and their correlation with the different abilities sought to measure. *Methods.* A framework for task definition and metric analysis is proposed. An explorative survey was first conducted with a board of experts to identify metrics to assess basic psychomotor skills. Based on the output of that survey, 3 novel tasks for surgical assessment were designed. Face and construct validation was performed, with focus on motion-related metrics. Tasks were performed by 42 participants (16 novices, 22 residents, and 4 experts). Movements of the laparoscopic instruments were registered with the TrEndo tracking system and analyzed. *Results.* Time, path length, and depth showed construct validity for all 3 tasks. Motion smoothness and idle time also showed validity for tasks involving bimanual coordination and tasks requiring a more tactical approach, respectively. Additionally, motion smoothness and average speed showed a high internal consistency, proving them to be the most task-independent of all the metrics analyzed. *Conclusion.* Motion metrics are complementary and valid for assessing basic psychomotor skills, and their relevance depends on the skill being evaluated. A larger clinical implementation, combined with quality performance information, will give more insight on the relevance of the results shown in this study.

Introduction

Over the past few decades, laparoscopic surgery has become a standard approach for several specialties, including gastrointestinal, gynecological, and urological surgery.¹ The benefits for the patient when using these techniques are evident; however, special surgical skills are required. These involve mainly adapting to the tactile and visual sensory limitations derived from indirect manipulation and visualization.² Training and assessment become thus a serious endeavor, as the acquisition of the necessary laparoscopic skills requires additional training, which often takes a long time.³

Simultaneously, social demands for safer surgeries and the need to minimize the learning time of future surgeons have led a change of paradigm in the training and accreditation of new professionals.⁴ Halsted-based

training⁵ is gradually being complemented by structured formative programs, based on standardized checklists and global rating scales (such as the Objective Structured Assessment of Technical Skills or OSATS) used to reduce bias on the final scoring.⁶ There is, however, a tendency to delay the moment the trainee is actively

involved in the operating room until he or she is deemed to have acquired basic laparoscopic skills. In consequence, the first stages of basic psychomotor skills, training take place in controlled laboratory settings by means of physical and virtual simulators.⁴ For accreditation purposes, however, no globally accepted training or assessment criteria exist.⁷ In several countries, standardized programs (such as the SAGES Fundamentals of Laparoscopic Surgery program (FLS) in the United States⁸ or the Dutch Cobra-alpha courses for gynecology⁹) are being pursued at the national level.

The development of tracking systems together with the advances on computing technology, has allowed refining the training and assessment techniques employed. Those include augmented physical simulators that make use of tracking devices such as ICSAD (Imperial College, London, UK),¹⁰ BlueDragon (University of Washington, Seattle, WA),¹¹ or TrEndo (Delft University of Technology, Delft, the Netherlands)¹²; or virtual reality simulators as MIST-VR (Mentice AB, Göteborg, Sweden),¹³ LapMentor (Simbionix, Lod, Israel),¹⁴ SIMENDO (Simendo, Rotterdam, the Netherlands),¹⁵ or SINERGIA.¹⁶ In these training systems, metrics used to assess tracked instrument motions are recurrently used. However, according to van Hove et al,¹⁷ reported levels of evidence on their validity are insufficient, consisting of nonconsecutive studies and/or lacking the application of adequate reference standards.¹⁸ Besides that, metrics are defined and interpreted in a general way, but more specificity with respect to the surgical skill to evaluate is needed.¹⁹ It is not clear which movement-related metrics measure specific surgical abilities. In consequence, evaluation of surgical performance in the training systems is being done based on different combinations of motion-related parameters, without taking into account their precise usefulness.²⁰

Within this context, the approach of this work is to analyze the needs and requirements of an environment for the assessment of surgical skills, focused on the tasks to use, the skills to assess, and the motion metrics employed to that end. Three main contributions arise from this global aim: (a) the development of a framework for assessment tasks' design from which we present and validate 3 new box trainer exercises; (b) the specification of new motion metrics such as "idle time" and an in-depth analysis of the influence of speed in skills assessment; and (c) based on validation results, the specification of motion metrics directly related to the skills intended to be assessed by a specific task.^{1,21} The framework consists of three phases: First, an explorative survey was conducted among a board of surgical experts to ascertain views regarding tasks, abilities, and important metrics required for assessment purposes. Based on that knowledge, in the second phase, 3 novel tasks were designed and built. The third phase—validation—was performed

by means of a questionnaire, tracking of the motions of the laparoscopic instruments, and statistical analysis of results.

Materials and Methods

Explorative Survey

Preliminary identification of the most relevant skills and the personal impressions of surgical experts regarding motion-related metrics was done by means of an explorative questionnaire. Fifteen expert surgeons with teaching experience at the Jesús Usón Minimally Invasive Surgery Centre (JUMISC, Cáceres, Spain) answered the questionnaire. The first section of the questionnaire comprised demographic and laparoscopic experience questions. Subsequent sections included general questions related to surgical assessment. In the last section of the questionnaire, participants were asked to give their views regarding the relevance of the basic psychomotor skills²¹ and the most prominent evaluation metrics found in the literature,^{19,22-25} rating them by means of Likert-type scale-based questions.

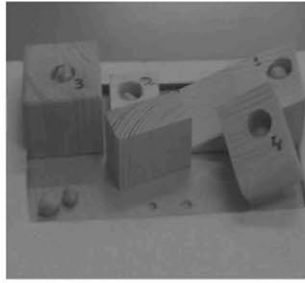
Task Design

Task design is a key aspect in any training and assessment program, and usually similar tasks are used throughout different physical and virtual simulators to train and assess a given ability.²⁶⁻²⁹ A design process of 3 new tasks for surgical assessment was carried out based on the results provided by the explorative survey. Supervision of this phase was performed by clinical experts both at the JUMISC and at the Leiden University Medical Centre (LUMC, Leiden, the Netherlands).

Clinical Validation

All clinical trials were carried out at the skills laboratories at the LUMC. A total of 42 participants took part in the experiments: 16 novices with no prior laparoscopic experience (N), 22 residents, out of whom 8 had a low laparoscopic experience (RL, less than 10 surgeries performed) and 14 a high experience (RH, 11-100 surgeries performed), and 4 experts (E, more than 100 surgeries performed). Three of the participants were left handed. As there is no gold standard for determining laparoscopic expertise, the threshold for forming the 4 categories was set based on the number of laparoscopic procedures performed by the participants.

For construct validation, all participants carried out the 3 tasks in the same order. Since the aim of the experiment was to validate the tasks for evaluation purposes, no prior trials were allowed. A brief explanation of the tasks'

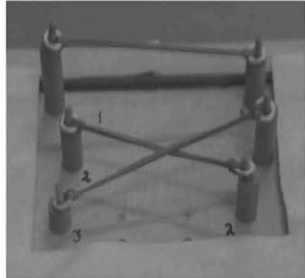


Grasp and Place

Peg grasping task, where a series of spherical objects (chickpeas) are to be placed on corresponding sequential holes. The scenario includes different blocks providing different heights, orientations and occlusion zones that add value to the assessment of spatial skills.

This task is performed with the **dominant hand only**.

Assessed skills: Hand-eye coordination, spatial perception, grasping.



Coordinated Pulling

Task that requires placing three elastic bands through their corresponding posts using solely two laparoscopic graspers. Pairs of posts vary regarding height and orientation with respect to camera (0, 135, 25 degrees respectively) within the setting. Force interactions are indirectly tested by the degree on which subjects manipulate the elastic bands.

Bimanual task.

Assessed skills: Hand-eye coordination, bi-manual coordination, grasping, haptic perception (indirectly).



Grasp and Transfer

Peg transfer task, where a small wooden cylinder is moved through a series of rings, distributed at different heights and angles. Transfer varies from left-to-right hand to right-to-left hand depending on the ring.

Bimanual task.

Assessed skills: Bi-manual coordination, spatial perception, grasping, suture (partially).

Figure 1. Clinical validation: (Left) Setting employed; (Right) Box trainer schematic disposition and TrEndo/metrics coordinate system

objectives was given to participants to let them infer, based on their own experience and skills, the best strategy to perform them. The clinical setting is shown in Figure 1. To ensure the same conditions for all participants, the position of the tasks within the box trainer, and the incision points for the camera and instruments were identical for each participant. Moreover, the starting and ending positions of the tips of the instruments and the order of targets placement at predefined positions (indicated by the numbers located next to them) were fixed for each task and were the same for all the participants. The endoscope provided a 0° view on a monitor, and was fixed at the same angle for all tasks and participants. Neither were time restrictions imposed nor was completion forced upon.

The TrEndo tracking system was employed for registering motions of laparoscopic instruments.¹² The following 10 parameters have been derived from the data collected with the TrEndo:

1. Time (T): Total time taken to complete the task(s).
2. Path length (PL): Total path followed by the tip of the laparoscopic instrument (m).
3. Depth (D): Total distance traveled by the instrument along its axis (m).
4. Motion smoothness (MS): Changes in acceleration, which reflect jerky movements of the surgical instrument (m/s^3).
5. Angular area (AA): Area between the farthest positions occupied by the tip of the instrument while performing a task (rad^2).
6. Volume (V): Volume of the 3-dimensional ellipsoid plotted around the standard deviation of the motions in an instrument (m).
7. Average speed (S, S_x , S_y , S_z): Rate of change of the instruments' tip position per second. Results are measured for the total magnitude (S) and in each Cartesian direction of the

box trainer; Sx (transversal), Sy (height), Sz (depth) (mm/s).

8. Maximum covered surface (Su): Maximum area covered by the instrument in the task surface plane, measured at the tip of the instrument (Figure 1): $[\text{Max}(x) - \text{Min}(x)] \times [\text{Max}(z) - \text{Min}(z)]$ (cm²).
9. Maximum volume (MV): Maximum cubic volume covered by the laparoscopic instrument within the box trainer measured at the tip of the instrument (Figure 1): $\text{Su} \times [\text{Max}(y) - \text{Min}(y)]$ (cm³).
10. Idle time (IT): Percentage of time during which the laparoscopic instrument was held still. A threshold was preestablished at a speed less than 5 mm/s (%).

The metrics explored in this study were chosen following previous studies (parameters 1-6) or derived from them (parameters 7-9)^{7,19,22} and from the explorative survey (parameter 10). The first 6 parameters were pre-calculated by the TrEndo 1.0 software, whereas the rest of the metrics were extracted from measurements using Matlab Release 2009b (Mathworks, Natick, MA).

Since validation studies featuring speed of motions as a metric are scarce,^{22,30,31} we wanted to investigate in detail its usefulness as a motion metric. A detailed analysis of its influence in all of space's directions was performed, similar to the decomposition performed by Megali et al,³² who analyzed the influence of acceleration. Both force interaction and quality metrics were discarded as they fell outside of the scope of this study. Other metrics, such as economy of movements and path deviation, were excluded because, among others, there is no clear definition that would indicate what the optimal movements in laparoscopy are.³³

For face validation, each participant was asked to fill out a questionnaire at the end of the trials. The questionnaire included a demographic survey, questions for task evaluation, and additional space for free comments. Participants were asked to rate the tasks performed according to (a) how they perceived the usefulness of the task for the assessment of psychomotor skills, (b) the design and construction of the task, (c) the difficulty on the execution of the task, and (d) an overall impression. Likert-type scales were used to rate each parameter.

Statistical Analysis

Statistical significance between subject groups was tested using Mann-Whitney test. Significant differences were considered at $P < .05$ both for face and construct validations. SPSS version 17 (SPSS Inc, Chicago, IL) was used for all statistical analysis. Construct validation

was measured at 2 sensitivity levels: lower and higher. The lower level of sensitivity was sought among experienced (Ex, comprising RH and E) and nonexperienced (NEx, comprising N and RL) participants. The higher level of sensitivity was established for the 4 predefined experience groups. For analytical purposes, 2-handed task metrics were analyzed separately, considering the dominant (D) and the nondominant hand (ND).

A novel aspect in this study dealt with the correlation of motion metrics and psychomotor skills. In a previous work, we sought to relate tasks and skills found in the literature with the metrics validated for each of them.³⁴ Following that study, we performed an analysis cross-referencing valid parameters for each task and psychomotor skills to relate the metrics explored with the abilities assessed by the tasks. For this analysis, we considered a metric valid if it presents significant differences between Ex and NEx and/or at least 2 groups from the N, RL, RH, and E populations.

Cronbach's α test for standardized items was performed to analyze metrics' reliability, as a means to measure the internal consistency (or reliability) between metrics for the different tasks.³⁵ A high value of reliability ($\alpha > .7$) implies that the metric is task-independent, and therefore useful regardless of the task's goals.²⁴ A lower value of α on the other hand indicates a higher degree of dependence of a metric to a given task.

Results

Explorative Survey

Out of 15 experts, 14 completed all questions of the survey, whereas one of them left out all questions regarding relevance of metrics. The results of the explorative survey showed a positive tendency of surgical experts toward the skills and metrics analyzed (Table 1). Hand-eye coordination obtained the highest scores (87% experts considered it relevant). Other important identified metrics were spatial perception (80% experts considered it relevant), bimanual coordination and dissection (73% experts considered it relevant). Haptic perception was rated as the least relevant skill (only 33% considered it relevant).

On the whole, all metrics proposed obtained a high median score. Individually highest rated metrics were depth, economy of movements and end-result analysis (56% considered them relevant), motion smoothness and forces interaction (49% considered them relevant). Metrics such as total and partial times, camera depth movements, speed, path length, and path deviation also obtained good results. Idle state was the lowest rated skill. Participants also indicated that, from the clinical point of view, it is important to assess the laparoscopic

Table 1. Explorative Survey Results^a

Question: Grade each of these psychomotor abilities according to their relevance for surgical evaluation of expertise (1 = *not relevant*, 3 = *mildly relevant*, 5 = *very relevant*)

Psychomotor skills	Definition	Score
Haptic perception	Tactile information perceived by the surgeon via the instruments	4 (2-5)
Camera navigation	Handling and orientation of the endoscope	5 (4-5)
Hand–eye coordination	Coordination between the inverted movements (due to the fulcrum effect of the trocar) of the instruments and the visual information presented in the monitor	5 (4-5)
Suture	Suture and knotting skills	4 (3-5)
Spatial perception	Dominion of depth and distances within the surgical workspace	5 (4-5)
Bimanual coordination	Ability to manipulate and transfer objects with two instruments simultaneously	5 (4-5)
Grasping	Steady and delicate manipulation of objects within the surgical workspace	5 (3-5)
Dissection	Cutting and dissection abilities	5 (4-5)

Question: Grade each of these metrics according to their relevance as measurements of surgical psychomotor skills (1 = *not relevant*, 3 = *mildly relevant*, 5 = *very relevant*)

Metrics	Definition	Score
Path length	Total path followed by the laparoscopic instrument	4 (2-5)
Path deviation	Ratio between the ideal path for a task and the total path length	4 (3-5)
Economy of movements	Ratio between the shortest distance necessary to complete a task and the total path length	5 (2-5)
Idle states	Number of periods when instrument movements/interactions are minimal	3 (2-5)
Depth	Total path length traveled by the instrument in the axis direction	5 (3-5)
Camera depth movements	Total path length traveled by the camera in the axis direction	4 (3-5)
Total time	Total time to perform a task	4 (3-5)
Partial time	Time taken to perform each of the steps of a task/exercise	4 (3-5)
Idle time	Time periods when instrument movements/interactions are minimal	4 (1-5)
Speed	Rate of change of the instruments' position per second	4 (2-5)
Acceleration	Rate of changes in the instruments' speed per second	4 (2-5)
Motion smoothness	Changes in acceleration, which reflect jerky movements of the surgical instrument	4.5 (1-5)
Force interactions	Instrument–tissue force and torsion interactions	4 (3-5)
End result analysis	Final score of the task performed	5 (4-5)
Sequence of actions	Measurement of the correct order of actions to perform for a given task/exercise	4 (3-5)
Repetition of actions	Number of repetitions required on a task before achieving satisfactory completion	4 (1-5)
Instrument presence	Total time were the instruments fall outside the camera's view	4 (1-5)

^aSurgeons were asked to grade the relevance of a series of psychomotor skills and assessment metrics. All data are expressed in terms of median (minimum–maximum).

competence based on both motion-related metrics and quality metrics (such as end-result analysis, errors performed), since assessment based solely on motion-related metrics does not adequately verify laparoscopic skills.

Regarding deviation from the optimal path, an additional question aimed to determine the way surgeons perform movements was posed, offering the alternatives presented by Chmarra et al³³ in their study. In this sense, no consensus was reached: 47% of the experts considered that the movement follows a circular path provoked by the rotation of surgical instrument on the incision point, 33% that it follows a straight line from one point to the other, and 20% that there is a retracting–seeking movement involved.

Task Design

The results of the explorative survey were used as guideline in the design process. Three tasks addressing directly most of the skills indicated in the survey were built: *Grasp and Place*, *Coordinated Pulling*, and *Grasp and Transfer* (Figure 2). The number of tasks was chosen as to provide a significant variety of exercises covering the basic laparoscopic skills.²¹ Skills that were not addressed included camera navigation, dissection and suture, although in the *Grasp and Transfer* task, a number of skills needed to perform cutting and/or suturing are already covered.

One of the aspects we sought was to introduce novelty with respect to the tasks identified for existing physical or

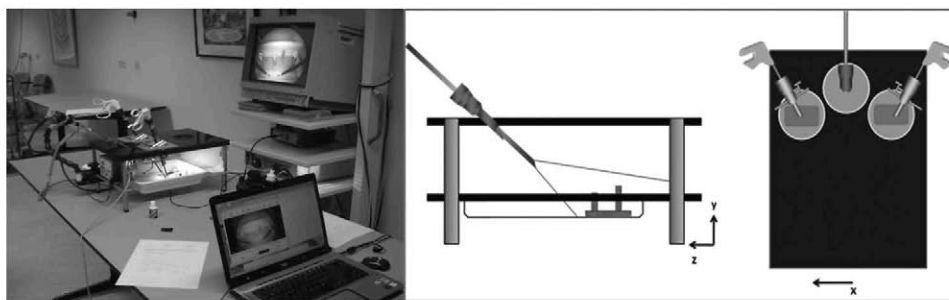


Figure 2. Task description and skills assessed

virtual simulators. Therefore, in the *Grasp and Place* task, we varied heights and orientations for the target holes, and included obstacles and partial occlusions. In the *Coordinated Pulling* task we made a variation of the height of the posts and in the orientation required to place the rubber band through the posts. The task required from the participants simultaneous bimanual coordination of the movements and forces because of the use of the elastic bands that involved applying pulling forces. The applying forces were lower than 4 N, and thus comparable with the forces used in laparoscopic surgery.³⁶ The novel element in the *Grasp and Transfer* task was introduced by placing the posts at different heights and angles. Additionally, 2 of the loopholes required left-to-right instrument peg transfer, whereas the other two loopholes required right-to-left transfer.

Clinical Validation

Construct validation. Validation results for all the 3 tasks are presented in Table 2 and Figures 3 to 5. The order of task performance was established in what was deemed to be increasing difficulty: *Grasp and Place*, *Coordinated Pulling*, and *Grasp and Transfer*. In general, nearly all participants were able to complete the 3 tasks. Only 2 participants were unable to finish 1 of the 3 tasks (one resident with low level (RH) of expertise failed to complete *Coordinated Pulling*, and one novice (N) did not complete *Grasp and Transfer*). Since this study focused on motion aspects rather than on qualitative results, we decided not to exclude the performance data of those 2 participants. Additionally, one RL and one RH did not perform the *Grasp and Transfer* task.

Construct validation of the *Grasp and Place* task presents significant differences for time, path length, depth, average speed and idle time. In general, the task showed the fewest differences between groups: the greatest

differences in performance were found between novices and every other group. No significant differences were found between residents and experts.

Coordinated pulling has construct validation for time, path length, depth, and average speed (directions S, Sy, and Sz). Differences were more prominent for the non-dominant hand in the cases of motion smoothness and average speed. On the whole, the performance of novices was significantly worse than that of residents with high experience and experts.

Construct validation of the *Grasp and Transfer* task showed that almost all metrics reflected significant differences between groups (time and average speed showing a higher sensitivity). This was especially observed for the nondominant hand (motion smoothness, average speed, maximum surface, and idle time). Overall, differences between groups were more prominent for this task than any other: not only between novices and residents/experts but also between residents (high/low experience) and experts. For this task, we observed that the learning curve tended to be short; the participants spent on average 30% of the time on learning how to transfer the peg through the first ring, whereas peg transition through the 3 remaining rings took 70% of the remaining time.

The cross-reference study (Table 3) reveals that for all skills studied, time, path length, and depth were valid parameters regardless of the task's goal. Total average speed is shown to be a significant metric for hand-eye coordination and bimanual coordination tasks. Motion smoothness features as significant for the nondominant hand in both bimanual tasks. Finally, a correlation between idle times and spatial perception and bimanual coordination tasks can be identified.

Analysis of Cronbach's α showed that angular area, volume, maximum surface, and idle time presented the lowest reliability (Table 2), thus being considered more

Table 2. Construct Validation and Reliability Results^a

Metrics	Significance (2 Groups)			Significance (4 Groups)			Reliability α
	1. GP	2. CP	3. GT	1. GP	2. CP	3. GT	
Time	x	x	x	N-RL, N-RH, N-E	N-RH, N-E, RL-RH	N-E, RL-E, RH-E	.648
Dominant hand	1. GP	2. CP	3. GT	1. GP	2. CP	3. GT	
Path length	x	x		N-RH, N-E	N-RH, N-E	N-E, RH-E	.607
Depth	x	x		N-E	N-RH	N-E, RH-E	.495
Motion smoothness							.730
Angular area				N-RL, N-RH		N-E, RL-E, RH-E	.275
Volume						N-E	.000
Average speed	x	x	x	N-RH, N-E	N-RH		.631
Average speed: x-axis	x			N-RH, N-E			.432
Average speed: y-axis	x	x	x	N-RH, N-E	N-RH, N-E	N-E	.712
Average speed: z-axis	x	x	x	N-RH			.591
Maximum surface				N-RL			.337
Maximum volume							.369
Idle time	x			N-RH			.253
Nondominant hand	—	2. CP	3. GT	—	2. CP	3. GT	
Path length		x			N-RH	N-E, RH-E	.535
Depth		x			N-RH	N-E, RH-E	.517
Motion smoothness		x	x			N-RH, RL-RH	.802
Angular area							.380
Volume							.000
Average speed		x	x		N-RH, N-E	N-RL, N-RH, N-E, RL-E	.768
Average speed: x-axis			x		N-E	N-RL, N-RH, N-E	.416
Average speed: y-axis		x	x		N-RH, N-E	N-RL, N-RH	.804
Average speed: z-axis		x	x		N-RH, N-E	N-RH, RL-E, RH-E	.750
Maximum surface			x			N-RL	.174
Maximum volume							.321
Idle time						N-RL, N-E	.420

Abbreviations: GP, grasp and place; CP, coordinated pulling; GT, grasp and transfer.

^aSignificant metrics between experienced (Ex) and nonexperienced participants are marked as x. Significant metrics between novices (N), residents with low experience (RL), residents with high experience (RH), and experts (E) are given by pairs of groups. Reliability values with $\alpha > .7$ are given in boldface.

task dependent than the rest. Only motion smoothness, average speed for the dominant hand (y [height] direction) and average speed for the nondominant hand (total and for the y [height] and z [depth] directions) showed reliability fulfillment. Other metrics such as time, depth, path length for the dominant hand, and average speed for the dominant hand presented moderate values of reliability, without reaching the preestablished thresholds.

Face validation. In general, the participants' subjective evaluation of the tasks showed positive results (Table 4). The 3 tasks obtained a median grading of 4 in perceived usefulness, design, and overall grading. The highest scores per groups for perceived usefulness were given to *Coordinated Pulling* task, which was also considered to have the best design for an assessment task. The *Grasp and Transfer* task obtained the lowest results for design, especially among residents with high experience and

experts. On overall grading, however, it obtained the highest scores. The difficulty of the tasks showed an increasing curve: Significant differences in scoring were identified between novices and experts. *Grasp and Place* was considered to be the easiest one and *Grasp and Transfer* to be the most difficult one.

Discussion

This work is but a step in a more ambitious research plan in which we are analyzing the needs for an assessment system based on visual tracking of the tools, without using sensor technology.³⁷ In this article, our aim is to establish the requirements of motion metrics and their relationship to the different psychomotor skills. Results show the importance of said metrics in the assessment of psychomotor surgical skills for laparoscopic surgery. The

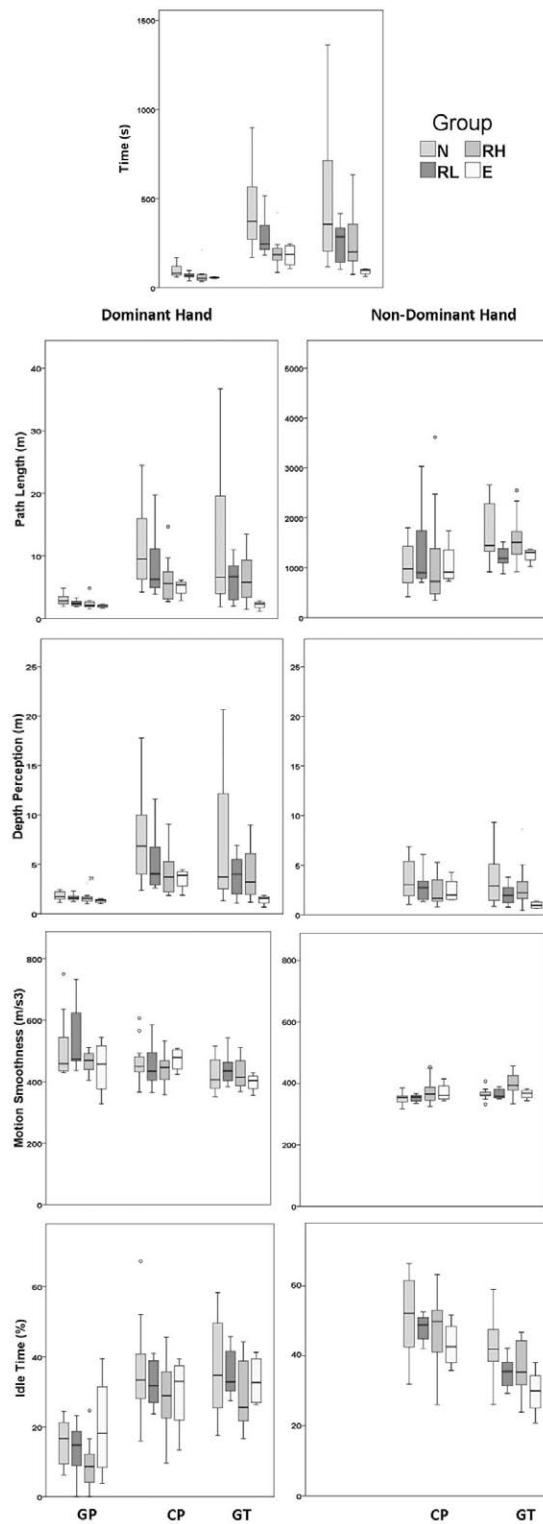


Figure 3. Construct validation: Scores for time, path length, depth, motion smoothness, and idle time

Results are presented as box and whisker plots, in which boxes represent interquartile range, bars represent medians, and whiskers represent the range of excluding outliers. Circles represent outliers. GP, task Grasp and Place; CP, task Coordinated Pulling; GT, task Grasp and Transfer.

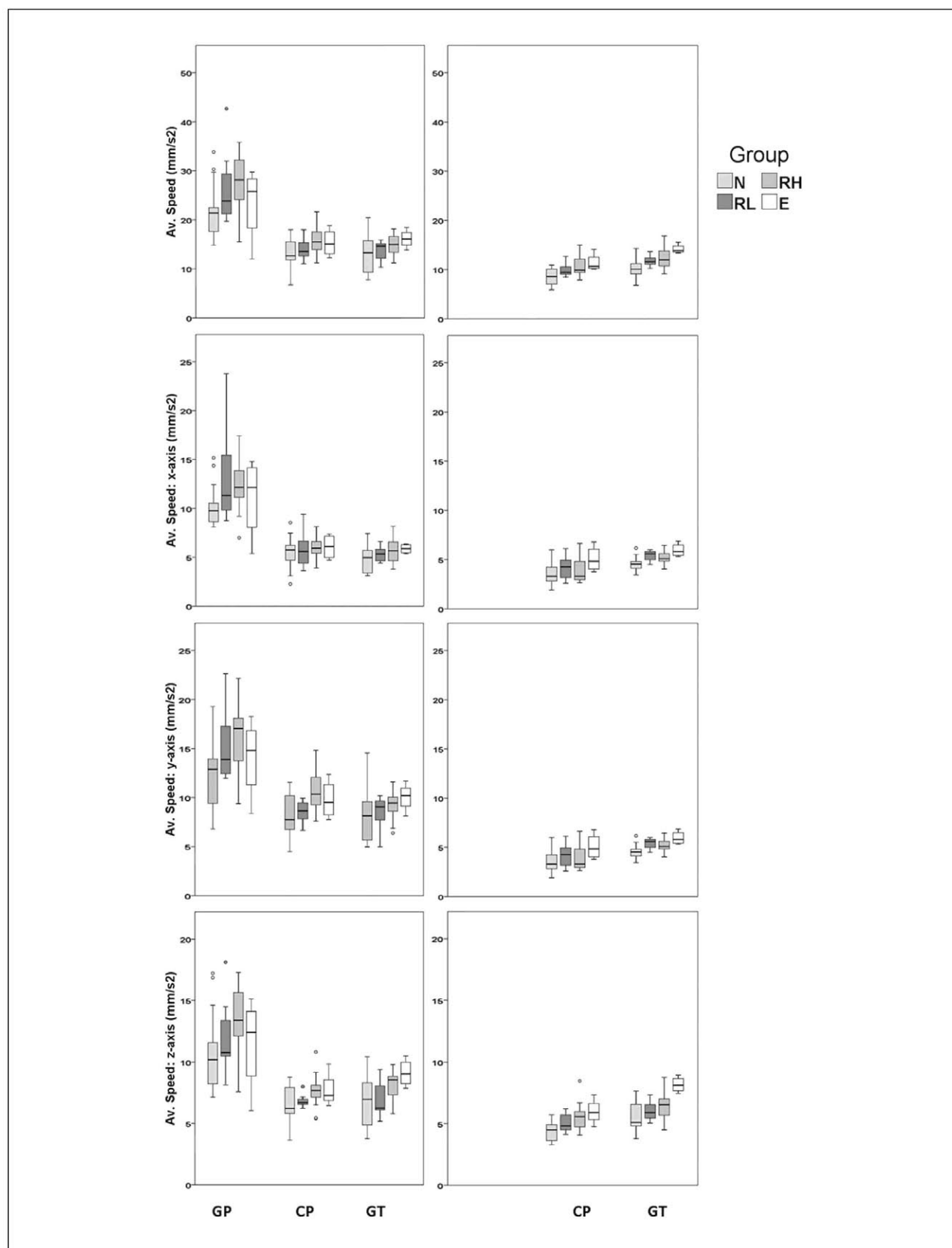


Figure 4. Construct validation: Scores for average speed metrics

Results are presented as box and whisker plots, in which boxes represent interquartile range, bars represent medians, and whiskers represent the range of excluding outliers. Circles represent outliers. GP, task Grasp and Place; CP, task Coordinated Pulling; GT, task Grasp and Transfer.

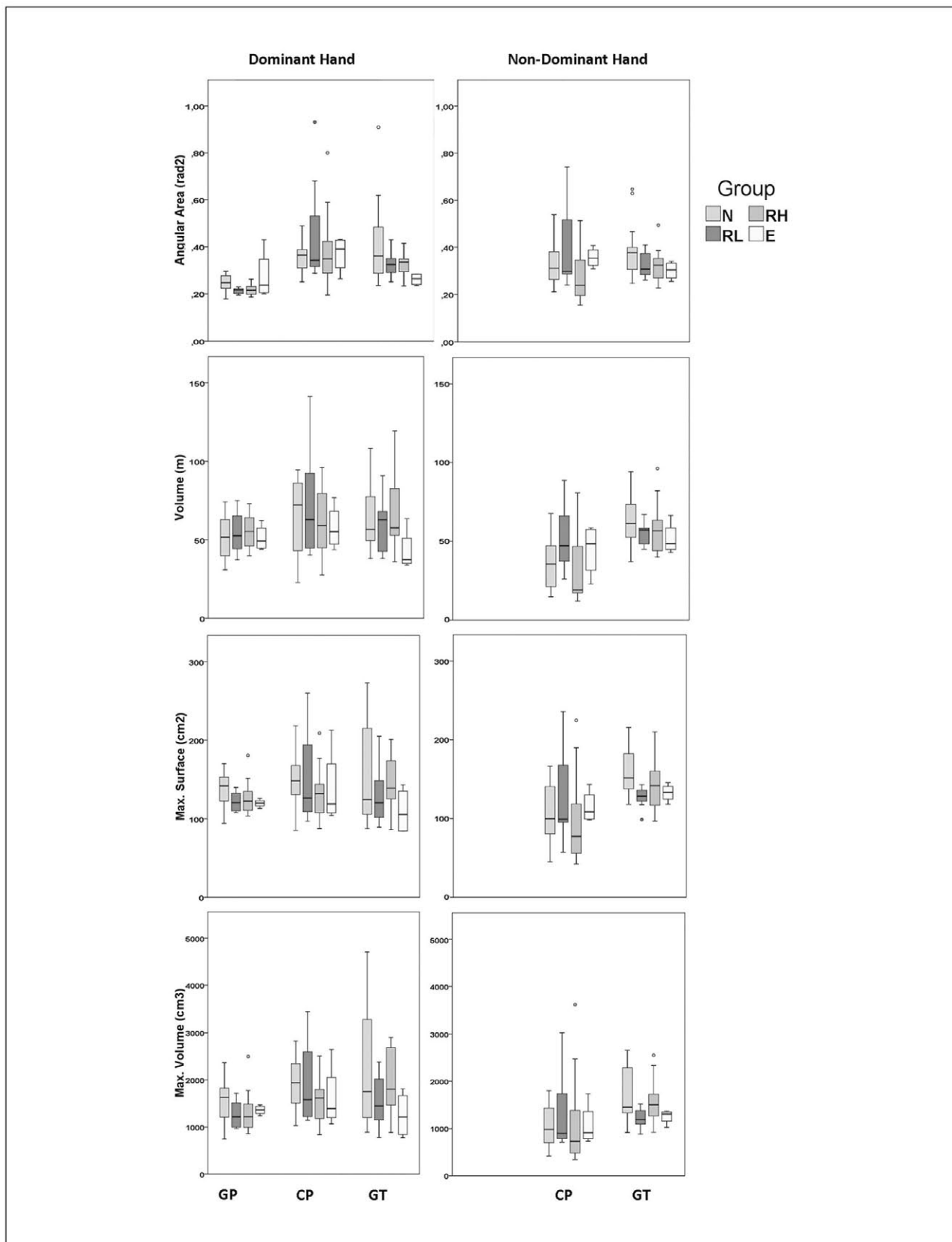


Figure 5. Construct validation: Scores for space-orientation metrics (angular area, volume, maximum surface, maximum volume) Results are presented as box and whisker plots, in which boxes represent interquartile range, bars represent medians, and whiskers represent the range of excluding outliers. Circles represent outliers. GP, task Grasp and Place; CP, task Coordinated Pulling; GT, task Grasp and Transfer.

Table 3. Cross-Reference Between Valid Metrics and Skills Assessed^a

	Hand-Eye Coordination (2)	Spatial Perception (2)	Bimanual Coordination (2)	Grasping (3)	Haptic Perception ^b (1)	Dissection ^b (1)	Suture ^b (1)
Time	2	2	2	3	1	1	1
Path length	2	2	2	3	1	1	1
Depth	2	2	2	3	1	1	1
Motion smoothness ^c	1	1	2	2	1	1	1
Angular area ^d	1	2	1	2		1	1
Volume ^c			1	1	1		
Average speed	2	1	1	2	1		
Maximum surface							
Maximum volume							
Idle time	1	2	2	2		1	1

^aThe score reflects the number of times the metric is found valid for that skill in particular (significance-2 groups and/or significance-4 groups for at least 2 pairs of groups), according to the relationship between tasks and skills featured in Figure 1. The number beside each skill (within parentheses) shows the maximum score that skill can obtain for the given tasks. Scores >1 are considered to be of interest.

^bIndirect abilities sought.

^cScores obtained solely for the nondominant hand.

^dScores obtained for the dominant hand.

Table 4. Face validation^a

	Total Median	N	RL	RH	E	P < .05
Perceived usefulness of the task for skills' assessment (1 = <i>little/no use</i> , 5 = <i>very useful</i>)						
Grasp and place	4	4 (2-5)	4 (2-5)	4 (3-4)	3.5 (3-5)	
Coordinated pulling	4	4 (3-5)	4 (3-5)	4 (2-5)	5 (4-5)	N-RH
Grasp and transfer	4	5 (2-5)	5 (3-5)	4 (3-5)	4 (3-5)	N-RH
Design and construction of the task (1 = <i>very poor</i> , 5 = <i>very good</i>)						
Grasp and place	4	4.5 (4-5)	4 (4-5)	4 (3-5)	4 (3-4)	N-E
Coordinated pulling	4	4.5 (2-5)	4 (2-5)	4 (2-5)	4 (3-5)	
Grasp and transfer	4	4.5 (3-5)	4 (2-5)	4 (1-5)	3.5 (3-4)	N-RL, N-RH, N-E
Difficulty of the task (1 = <i>very easy</i> , 5 = <i>very difficult</i>)						
Grasp and place	2	2 (1-3)	2.5 (1-5)	2 (1-4)	2 (1-3)	
Coordinated pulling	4	4 (2-5)	4 (3-5)	3 (2-5)	3.5 (3-5)	
Grasp and transfer	4.5	5 (3-5)	5 (3-5)	5 (2-5)	3 (2-3)	N-E
Overall grade of the task (1 = <i>very poor</i> , 5 = <i>very good</i>)						
Grasp and place	4	4 (3-5)	4 (3-5)	4 (2-4)	3 (3-4)	N-E
Coordinated pulling	4	4 (3-5)	4 (3-5)	4 (3-4)	4 (4-5)	
Grasp and transfer	4	5 (3-5)	4 (3-5)	4 (2-5)	4 (3-4)	N-RL, N-RH

^aSignificant differences between N (novices), RL (residents with low experience), RH (residents with high experience), and E (experts) are expressed as pairs of groups. All data are expressed in terms of median (minimum-maximum).

relevance of the most commonly used assessment parameters, for example, time, path length and depth, has been corroborated. Other less often used metrics, as for example, speed of motion, have been analyzed, and proved to be effective when discerning between several surgical skills.

When performing validation studies, one cannot fully establish the value of a certain metric for the assessment of surgical skills without considering all surrounding conditions. These include factors such as task definition, clinical setting, relative difficulty, previous practice of the tasks, or number of participants in validation studies,

among others. To better understand them, we developed a 3-phase framework consisting of (a) an explorative survey amongst experts, (b) a task design process, and (c) a clinical validation.

The explorative survey allowed us to learn firsthand the impressions of teaching experts regarding important aspects of surgical assessment. The study was conceived as a broad overview on the main skills sought and metrics that can be registered. This useful information served as the basic guideline for the task design process and the selection of motion metrics for clinical validation. There

were several drawbacks that affected the final weight the survey had, as the limited sample population or a slight predisposition toward positive answers (as shown by the general high grades). Moreover, the questionnaire was built on previously defined hypotheses, which resulted on a number of Likert-type closed questions. Although this scheme is more appealing to participants, it may also lead them to responses they would otherwise not give.³⁸

The 3 tasks built were chosen after a design process where several other ideas were analyzed and discarded. Common criticism of participants to task design was directed to the materials employed, in particular to the use of chick-peas in the *Grasp and Place* task and wooden cylinders for the *Grasp and Transfer* task, as they proved to be slippery and difficult to grasp. Moreover, several participants criticized the fact that the camera was situated at a fixed position, which had effect on altering participants' vision of depth. There were 2 reasons for camera fixation: (a) to provide exactly the same conditions of performing the assessment tasks to all the participants and (b) to make fixed-camera recordings of the tasks to investigate possibilities of visual tracking of the tools.³⁷ For skills' training, these aspects can be considered a drawback, since it is desirable to use pegs made of materials that is adequate for that end (eg, the one that imitates the characteristics of the tissue). For assessment purposes, our study shows that experienced surgeons and residents required less time than novices to master the tasks, which reflects a higher ability on their part to adapt and overcome more challenging conditions.

The tasks were designed to be of increasing difficulty. The a posteriori validation of the tasks indicated that we succeeded in that; more significant differences between groups have been found for the tasks that were intended to be more difficult. This was also ratified by participants: according to their subjective opinion, the difficulty curve increased with each task.

Clinical validation of tasks and metrics allowed us to (a) determine the construct validity of the tasks for assessment purposes, (b) determine the degree of dependence of each metric to the different skills, and (c) determine the correlation between tasks and skills assessed. Overall, we found that time, path length, and depth were good discriminators for all abilities considered (Table 2). This confirms the general trend observed in validation studies from diverse sources.^{26,39-41} As expected, these metrics show moderately high consistency levels. Motion smoothness also showed significant differences between groups, although from the validation results it can be seen that its influence was more significant on grasping tasks involving bimanual coordination (Table 3).

Space-orientation metrics (angular area, volume, maximum surface, maximum volume) showed uneven results and a low internal consistency. Their validity as independent metrics, therefore, has not been proved. Significance of the space-oriented metrics could be

directly related with the difficulty of the task rather than to the abilities assessed. Further studies are needed to confirm or to reject this hypothesis.

Idle time showed significant differences for *Grasp and Place/Grasp and Transfer* tasks. Since both those tasks involved a tactical approach to the transportation of a peg, it seems possible that the metric assessed the amount of time taken for planning the next steps (strategy). In any case, the metric proved useful for measuring skills such as spatial perception, bi-manual coordination or grasping. This contrasts with the opinions given by experts in the explorative survey, who were not so convinced about its usefulness. Therefore, further studies are needed to investigate that matter.

Analysis of average speed showed promising results of its usefulness as an evaluation metric (Figure 4). Significant differences were found for all 3 tasks, and internal consistency was overall high, reflecting an important degree of task independence. Moreover, there seems to be a relationship between speed and hand-eye coordination/grasping skills. We observed that significance of the speed was lower on the *x*-axis (constant depth plane) compared with the *y*-axis and *z*-axis (variable depth planes). Although further research is required to find the explanation to this, it is possible that expert surgeons' perception of depth (and thus depth-related navigation of instruments) is better than those of less experienced participants.

Finally, this study reflects the importance of distinguishing dominant and nondominant hands during the process of skills' assessment, as shown in other studies.⁴³ On a general basis, in bimanual tasks performance of the nondominant hand discriminates between the 4 groups better than the performance of the dominant hand. Moreover, nondominant hand motion-related metrics showed the highest consistency values. Because of the low number of left-handed participants, we did not further explore the differences between performance of the right-handed and left-handed participants, as done by Grantcharov et al.⁴⁴

Conclusion

The need for new structured, objective training and accreditation programs for laparoscopy has motivated much research over the past few years, focusing both on clinical aspects²⁵ (development of structured curricula, development of laboratory training settings, etc) as well as on the adoption of new technologies for sensing and tracking,⁴⁵ virtual simulation,⁴⁶ and e-learning.⁴⁷ For psychomotor skill learning, these combined efforts have led to the development of assessment methods that can provide both formative and summative objective feedback to trainees, based on force and motion analysis. Validation of the training⁴⁸ and assessment^{17,20,34} systems shows that, although motion analysis can provide useful information, the significance of the individual motion-related

metrics may vary between studies, depending on the tasks employed and/or skills assessed.

Establishing valid motion-related metrics for assessing laparoscopic psychomotor skills should be a multidisciplinary effort that involves both clinicians (from application field), experts in human factors (usability and understanding of human capabilities) and engineers (methods and tools to be applied).⁴⁹ In this study, we followed a novel framework for task design and validation. Through the whole process, a multidisciplinary team worked on better understanding the influence motion-related metrics have on the assessment of laparoscopic skills and the determination of surgical expertise.

This study can serve as a guideline for a structured way of seeking and analyzing motion-related metrics for assessment purposes as well as developing new assessment tasks. Further insight on whether results obtained for the metrics are universal or specific for the tasks and skills proposed may be derived from a larger clinical implementation on training courses, as well as from the use of other existing tasks. Since evaluation of psychomotor skills should not be based only on efficiency parameters (both motion and forces), new studies should also take into account the influence that qualitative metrics (eg, errors, end-product information) have over the global processes of skill assessment.

Acknowledgments

The authors thank all participants and personnel involved both of the explorative survey (JUMISC, Cáceres) and the clinical trials (LUMC, Leiden). Authors also thank the valuable contributions of Francisco Gayá (Madrid) and Thom van Beek (Delft) during the tasks' construction process.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

This article and all related research are being partially funded by the FPU Program of the Spanish Ministry of Science and Innovation (AP2007-00465).

References

1. Usón J, Sánchez-Margallo FM, Pascual S, Climent S. *Formación en Cirugía Laparoscópica Paso a Paso*. 4th ed. Cáceres, Spain: Minimally Invasive Surgery Centre Jesús Usón; 2010.
2. Cuschieri A. Laparoscopic surgery: current status, issues and future developments. *Surgeon*. 2005;3:125-130.
3. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Ann Surg*. 2004;240:518-528.
4. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg*. 2004;91:1549-1558.
5. Halsted WS. The training of surgeons. *Johns Hopkins Hosp Bull*. 1904;15:267-275.
6. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ*. 1975;1:447-451.
7. Chmarra MK, Grimbergen CA, Jansen FW, Dankelman J. How to objectively classify residents based on their psychomotor laparoscopic skills? *Minim Invasive Ther Allied Technol*. 2010;19:2-11.
8. Ritter EM, Scott D. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov*. 2007;14:107-112.
9. Hiemstra E, Kolkman W, Jansen FW. Skills training in minimally invasive surgery in Dutch obstetrics and gynecology residency curriculum. *Gynecol Surg*. 2008;5:321-325.
10. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg*. 2001;193:479-485.
11. Rosen J, Brown JD, Barreca M, Chang L, Hannaford B, Sinanan M. The Blue DRAGON—a system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment. *Stud Health Technol Inform*. 2002;85:412-418.
12. Chmarra MK, Bakker NH, Grimbergen CA, Dankelman J. TrEndo, a device for tracking minimally invasive surgical instruments in training setups. *Sensor Actuat A Phys*. 2006;126:328-334.
13. Mentice. Improving patient safety. <http://www.mentice.com>. Accessed August 21, 2012.
14. Simbionix. LAP Mentor laparoscopic surgery simulator for general surgery, gynecology and urology. http://www.simbionix.com/LAP_Mentor.html. Accessed August 21, 2012.
15. Simendo. SIMENDO laparoscopy simulator and arthroscopy simulator. <http://www.simendo.eu>. Accessed August 21, 2012.
16. Lamata P, Gómez EJ, Sánchez-Margallo FM, et al. SIN-ERGIA laparoscopic virtual reality simulator: didactic design and technical development. *Comput Methods Programs Biomed*. 2007;85:273-283.
17. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97:972-987.
18. Phillips B, Ball C, Sackett D, et al. *Levels of Evidence and Grades of Recommendation*. Oxford, England: Oxford Centre for Evidence-based Medicine; 2001.
19. Cotin S, Stylopoulos N, Ottensmeyer M, et al. Metrics for laparoscopic skills trainers: the weakest link! In: Dohi T, Kikinis R, eds. *Medical Image Computing and*

Computer-Assisted Intervention—MICCAI 2002. Heidelberg, Germany: Springer; 2002:35-43.

20. Thijssen AS, Schijven MP. Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? *Am J Surg*. 2010;199:529-541.
21. Lamata P. *Methodologies for the Analysis, Design and Evaluation of Laparoscopic Surgical Simulators* [dissertation]. Louvain: Presses Universitaires de Louvain; 2006.
22. Cavallo F, Megali G, Sinigaglia S, Tonet O, Dario P. A biomechanical analysis of surgeon's gesture in a laparoscopic virtual scenario. *Stud Health Technol Inform*. 2006;119:79-84.
23. Rosen J, Hannaford B, Richards C, Sinanan M. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Trans Biomed Eng*. 2001;48:579-591.
24. Fried GM, Feldman LS. Objective assessment of technical performance. *World J Surg*. 2008;32:156-160.
25. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective assessment. *Surg Endosc*. 2003;17:220-226.
26. Eriksen JR, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality stimulator. *Surg Endosc*. 2005;19:1216-1219.
27. Gallagher A, Lederman A, McGlade K, Satava R, Smith C. Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance. *Surg Endosc*. 2004;18:660-665.
28. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. *J Am Coll Surg*. 2006;202:779-787.
29. Verdaasdonk EGG, Stassen LPS, Schijven MP, Dankelman J. Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. *Surg Endosc*. 2007;21:1406-1412.
30. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *Br J Surg*. 2009;96:1086-1093.
31. Sokollik C, Gross J, Buess G. New model for skills assessment and training progress in minimally invasive surgery. *Surg Endosc*. 2004;18:495-500.
32. Megali G, Sinigaglia S, Tonet O, Cavallo F, Dario P. Understanding expertise in surgical gesture by means of hidden Markov models. *Proc IEEE RAS EMBS Int Conf Biomed Robot Biomechatron*. 2006:625-630.
33. Chmarra MK, Jansen FW, Grimbergen CA, Dankelman J. Retracting and seeking movements during laparoscopic goal-oriented movements. Is the shortest path length optimal? *Surg Endosc*. 2008;22:943-949.
34. Oropesa I, Sánchez-González P, Lamata P, et al. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. *J Surg Res*. 2011;171:e81-e95.
35. Sánchez-Peralta LF, Sánchez-Margallo FM, Moyano-Cuevas JL, et al. Construct and face validity of SINERGIA laparoscopic virtual reality simulator. *Int J Comput Assist Radiol Surg*. 2010;5:307-315.
36. de Visser H, Heijnsdijk EAM, Herder JL, Pistecky PV. Forces and displacements in colon surgery. *Surg Endosc*. 2002;16:1426-1430.
37. Sánchez-González P, Cano AM, Oropesa I, et al. Laparoscopic video analysis for training and image-guided surgery. *Minim Invasive Ther Allied Technol*. 2011;20:311-320.
38. Williams A. How to write and analyse a questionnaire. *J Orthod*. 2003;30:245-252.
39. van Dongen KW, Tournioij E, van der Zee DC, Schijven MP, Broeders IAMJ. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc*. 2007;21:1413-1417.
40. Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM. Construct validity of the LapSim laparoscopic surgical simulator. *Am J Surg*. 2006;191:28-32.
41. van Sickle KR, McClusky DA 3rd, Gallagher AG, Smith CD. Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. *Surg Endosc*. 2005;19:1227-1231.
42. Yamaguchi S, Konishi K, Yasunaga T, et al. Construct validity for eye-hand coordination skill on a virtual reality laparoscopic surgical simulator. *Surg Endosc*. 2007;21:2253-2257.
43. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. *Surg Endosc*. 2003;17:1082-1085.
44. Chmarra MK, Grimbergen CA, Dankelman J. Systems for tracking minimally invasive surgical instruments. *Minim Invasive Ther Allied Technol*. 2007;16:328-340.
45. Oropesa I, Lamata P, Sánchez-González P, et al. Virtual reality simulators for objective evaluation on laparoscopic surgery: current trends and benefits. In Kim J, ed. *Virtual Reality*. Rijeka, Croatia: InTech; 2011.
46. Sánchez-González P, Oropesa I, Romero V, et al. TELMA: technology enhanced learning environment for minimally invasive surgery. *Proc Comput Sci*. 2011;3:316-321.
47. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg*. 2008;95:1088-1097.
48. Freudenthal A, Stüdeli T, Lamata P, Samset E. Collaborative co-design of emerging multi-technologies for surgery. *J Biomed Inform*. 2011;44:198-215.